1. INTRODUCTION

1.1. Cancer

Cancer, also known as malignant neoplasm or malignant tumor, is a group of diseases which can affect any part of the body by causing abnormal cell growth with capability to spread to the other parts of the body. In cancer a single cell or group of cells exhibit uncontrolled growth and devastation of the body tissue. The potential symptoms of cancer include a new lump with inflammation, prolonged cough, unusual or irregular bleeding alteration or disturbance in bowel movements and mysterious weight loss. This disease is caused because of many reasons such as, inappropriate diet, inappropriate exercise, various infections, and radiation; also the disease may be caused by hereditary factors, some physical and chemical agents, trauma or inflammation, or hormonal imbalance. The use of tobacco has been reported as the cause of around 22% of cancer deaths, whereas 10% of the deaths are due to obesity, lack of physical activity, poor diet and alcohol consumption.[1,17]

According to WHO report, 14.1 million new cases of cancer and 8.2 million cancer deaths were reported worldwide in 2012. More than 60% of world's total new annual cases occur in Africa, Asia, and Central and South America. These regions account for 70% of the deaths because of cancer in the world. According to the world cancer report 2014, the 5 most common sites of cancer diagnosed in men in 2012 were lung, prostate, colorectum, stomach, and liver cancer. Whereas in women the 5 most common sites detected were breast, colorectum, lung, cervix, and stomach cancer. According to the Department of Epidemiology and Biostatistics, Kidwai memorial institute of oncology, in India the estimated number of new cancers per year is about 7 lakhs and over 3.5 lakhs people die of cancer each year.[17,18.]

In a healthy person, cell completes its life cycle by passing through different phases which include, G_0 phase, actually the resting phase and S phase, wherein DNA synthesis occurs. Following the termination of DNA synthesis, cell cycle takes a small pause called G_2 phase, where various proteins check the DNA integrity and allow the cell to enter M phase. In M phase the cell passes through four different sub-phases namely, prophase, metaphase, anaphase and telophase. In M phase the cell containing double compliment of DNA divides into two daughter cells following cytokinesis. The damage or disturbance in any step of this normal cell cycle or

disturbances in cell cycle controlling genes or enzymes lead to abnormal cell division or aneuploidy and ultimately to cancer.[1,19]

At present many antineoplastic agents have been reported and many are under development. These agents act at different levels in cell cycle to control the cancerous condition. Majority of these agents come under the category of narrow therapeutic index and have greater potential for causing harmful side effects. Various drugs have been discovered which act at different stages of cell cycle. Despite having such a vast number of sites or targets for action there is no expected cure observed for cancer. This noncompliance is mainly either because of short therapeutic index of drugs or development of toxicity. As normal cells continuously proliferate in adult tissues it is expected that cytotoxic drugs cause varying degrees of tissue damage and thus cause toxicity to patients. The body also develops resistance to many drugs which limits the utility of such agents. The resistance arises because of loss of necessary activating enzymes, over-expression of drug efflux pump or mutation that raise the threshold for programmed cell death (apoptosis).[20]

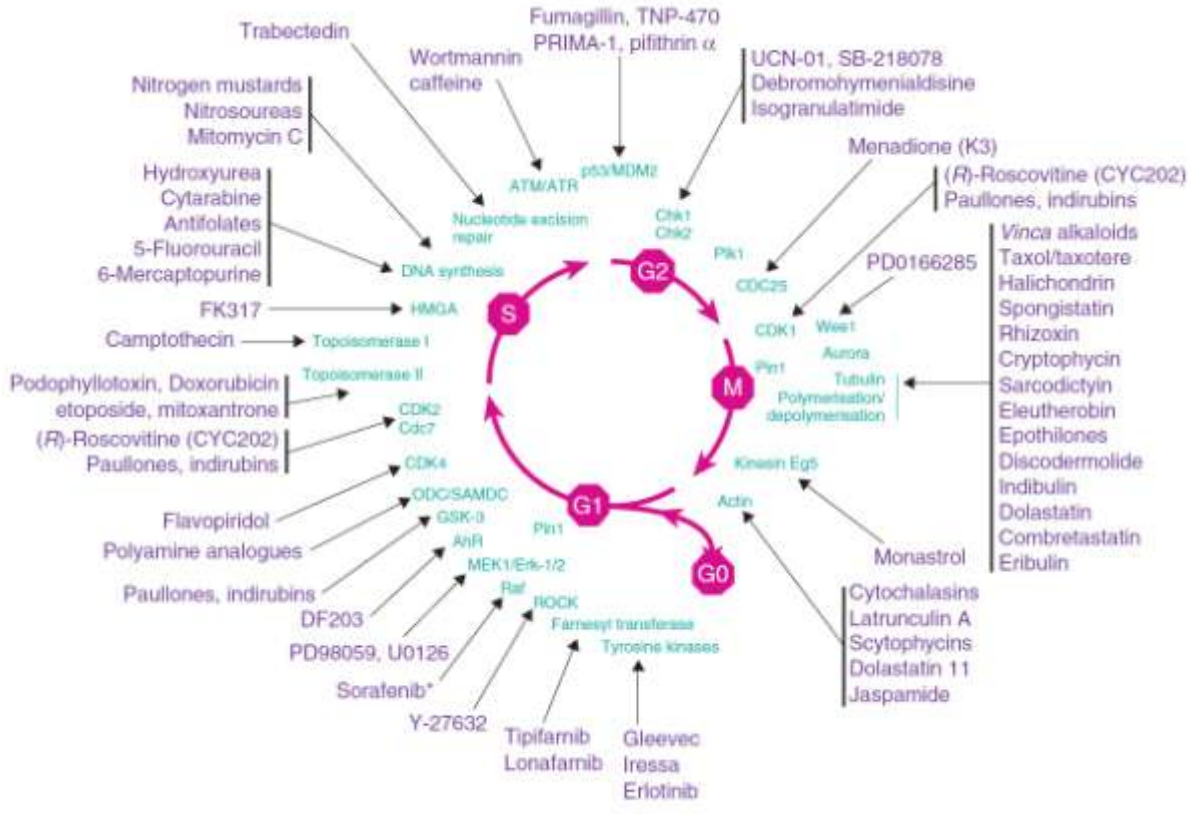

Figure 1: Different targets in cell cycle for cancer treatment with the drugs acting on them.[19]

In such conditions, functional genomic data is helpful for the identification of therapeutic target proteins which regulate the cell cycle and get expressed differentially in tumors compared with the normal cells of the adult tissues. It is conceivable that pharmaceutical targeting of such proteins would help in the development of a new generation of effective therapeutic agents that will have minimal host toxicity.[20]

Protein kinases are enzymes that modify other proteins by phosphorylation. This typically results in a functional change of the target protein i.e. substrate, by altering enzyme activity, cellular location or association with other proteins. There are more than 500 protein kinases which are associated with different physiological functions in the body. Kinase activity modifies almost up to 30% of all human proteins and are known to control the majority of cellular pathways, particularly those related to signal transduction. Protein kinases are mainly of four types, which include Tyrosine specific protein kinases, Serine/threonine specific protein kinases, Histidine specific protein kinases and mixed kinases. They phosphorylate the amino acids on which they are named in the cascade and are majorly associated with the signal transduction. Recent studies have demonstrated that various cell cycle regulating protein kinases show abnormally high levels of expression in tumors. They have been proposed as novel targets for the development of anticancer drugs.[4,5]

Among the various mitotic regulatory kinases, the aurora kinases belong to serine/ threonine kinases, which play vital role in the cell cycle regulation. It is observed that the family of aurora kinases is expressed at elevated levels in many human cancers. This kinase family is not only vitally important regulator of cell division but has also been shown to functionally interact with multiple critical oncoproteins and tumor suppressor proteins. Aurora kinase family is serine/threonine kinases which phosphorylate the hydroxyl group of serine or threonine. These kinases are essential for genetic material alignment, dispersion and cytokinasis during mitosis. Aurora kinase family is divided in to three sub classes, aurora A kinase, aurora B kinase and aurora C kinase.[6,7]

Aurora A localizes mainly to the centrosomes, late in the S phase and early in the G1 phase and expressed highest at the G2/M phase of the cell cycle, and is a key to correct centrosome maturation and separation. Aurora A also shows allegiance in chromosome alignment at metaphase plate. The high expression of Aurora A in mammalian cells leads to centrosome

amplification and polyploidy.[13,14,21] Aurora B localizes predominantly to the spindle midzones and functions in the attachment of the mitotic spindle to the kinetochore of the centrosome. It controls chromosome segregation and cytokinasis. Initially aurora B is required for the phosphorylation of histone H3 at serine 10. It then maintains a wait-anaphase signal until all chromosomes are in the correct orientation for separation. Aurora C is localized to Chr19q13 and was first isolated from a testis cDNA library. The role of aurora-C in tumorigenesis is less well-defined and understood to be similar to aurora B.[4,22]

In many studies the over-expression of aurora A and B in different tumors is correlated with cancer. Aurora kinases promote tumorigenesis not only by excessive phosphorylation of the cell's normal physiologic substrate but also by aberrant phosphorylation of cytoplasmic proteins which provoke genetic instability that leads to tumorigenesis. Thus inhibition of any of the isoforms of aurora kinases has shown cancer cell death by apoptosis and mitotic catastrophe.[6]

The crystal structure of aurora kinase shows that each member of the family consists of approximately 400 residues and has a conserved C-terminal catalytic domain and a short N-terminal domain that varies in size. These kinases are activated by phosphorylation of threonine (Thr288)/serine (Ser53/Ser342) residue. In normal physiology, ATP activates the aurora kinase. So it has been recognized that highly specific ATP competitive inhibitors can be obtained as cancer therapeutic agents. Fancelli *et al.* identified ATP binding pocket of the aurora A kinase. It mainly consists of kinase hinge region, buried region, phosphate binding region, sugar region and solvent accessible region (**Figure 2** represents ATP binding site with the nucleus of a general inhibitor nucleus). The buried region is small; the phosphate binding region is where the ATP tail is placed. The hinge region (residue 210 - 216) has an important role in forming the catalytic active site. In this hinge region the scaffold forms the hydrogen bonding mainly with Glu211 and Ala213 along with Lys162, Leu139 and Leu263.[23]

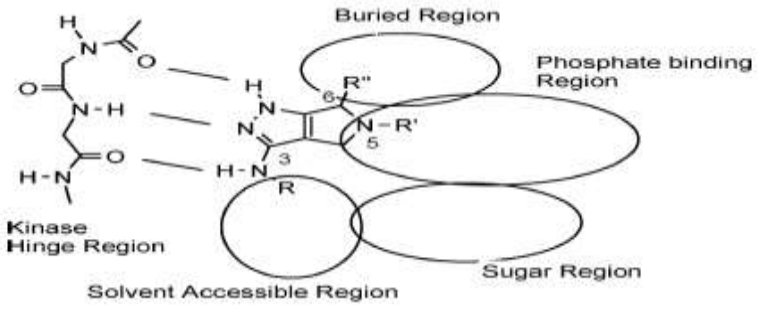

Figure 2: Schematic representation of ATP binding site with an inhibitor in aurora kinase.

Since the first clinical trial for aurora kinase inhibitor started in 2004, to date, more than ten small molecule inhibitors of these kinases have entered in clinical trials as potential anticancer agents, based on the ATP competitive inhibitor hypothesis. And these preclinical and preliminary results of clinical trials suggest that this class of agents is promising for cancer treatment.[24]

1.2. Computer aided drug design

It is well know that the traditional method of getting a drug with good safety is time consuming and costly. To minimize the money and time so many evolutionary techniques are introduced. Computer Aided Drug Design (CADD), also known as rational drug design/ computer aided molecular modeling/*in silico* drug design, is one such evolutionary technique. Computer based designing of molecules is robust, reliable and a rapid method that boosts the process of drug development and also reduces the uncertainty of process, huge random use of chemicals and of animals for testing the new chemical entities. Here in **section I** of the thesis, the design and systematic validation of two different computational models developed anti-cancer agents have been reported. The focus in cancer is on development of computational models on aurora kinase inhibitiors. Different tools used for the development of computational 3D and 4D QSAR models include AutoDock, Tripos Sybyl, OpenBabel, GROMACS, MATLAB, Schrodinger, Gaussian, PRODRG server and LQTA-QSAR tool. General basics of computational chemistry are described in the following section.

Computational chemistry is a branch of chemistry that uses computer simulations in solving chemical problems and designing newer chemical molecules for the desired biological targets. CADD involves diverse disciplinary aspects of physics and chemistry. Different computational chemistry tools fall within the domain of rational drug designing. Recent technological advances like structure based drug design, ligand based drug design and availability of large number of chemical and biological databases have made this technique more useful and interesting. So, the drug design and discovery process is a multidisciplinary team effort where computational chemists play a central role.[25]

Ligand based drug design (LBDD) (or indirect drug design) involves the concept of similarity searching, pharmacophore searching and quantitative structure activity relationship (QSAR). Ligand-based drug design relies on the knowledge of other molecules that bind to the

biological target of interest. These other molecules may be used to derive a pharmacophore model that defines the minimum necessary structural characteristics a molecule must possess in order to bind to the target. In other words, a model of the biological target may be built based on the knowledge of what binds to it, and this model in turn may be used to design new molecular entities that interact with the target. Alternatively, in QSAR a correlation between calculated properties of the molecules and their experimentally determined biological activities are derived. These QSAR relationships in turn may be used to predict the activity of new analogs.[26]

Structure based drug design (SBDD) (or direct drug design) involves the concept of protein-ligand docking i.e. molecular docking. Structure-based drug design relies on the knowledge of the three dimensional structure of the biological target obtained through methods such as X-ray crystallography or NMR spectroscopy. If an experimental structure of a target is not available, it is possible to create a homology model of the target on the basis of the experimental structure of a related protein(s). Using the structure of the biological target, candidate drugs that are predicted to bind with high affinity and selectivity to the target may be designed using interactive graphics and the intuition of a medicinal chemist.[27]

Current methods for structure-based drug design can be divided roughly into two categories. The first category is about "finding" ligands for a given receptor, which is usually referred to as database searching. The key advantage of database searching is that it saves synthetic efforts to obtain new lead compounds. Another category of structure-based drug design methods is about "building" ligands, which is usually referred as receptor-based drug design. In this case, ligand molecules are built up within the constraints of the binding pocket by assembling small pieces in a stepwise manner.[28,29] The general process of drug design/drug discovery is outlined in **Figure 3**.

The general procedure of drug discovery or drug design starts from the target identification. Target is nothing but the factor responsible for particular disease/disorder for which drug is to be discovered or designed. The lead identification for identified target mainly includes two types of computational approaches, i.e. SBDD and LBDD. The techniques used in these two approaches are summarized in the following sections.

1.2.1. Protein crystallography and homology modeling

Protein structure determination mainly involves the use of X-ray technique or NMR technique. X-ray crystallography is a very high resolution microscopy. It helps to envision the protein structures at the atomic level. Using this technique one can learn how the proteins interact with other molecules and their conformational changes. By using this type of information one can design novel drugs targeting a particular protein. The X-rays are diffracted by the electrons present in the structure and thus it results into a 3-dimensional map presenting the distribution of electrons in the structure. In order to perform this experiment, the isolated

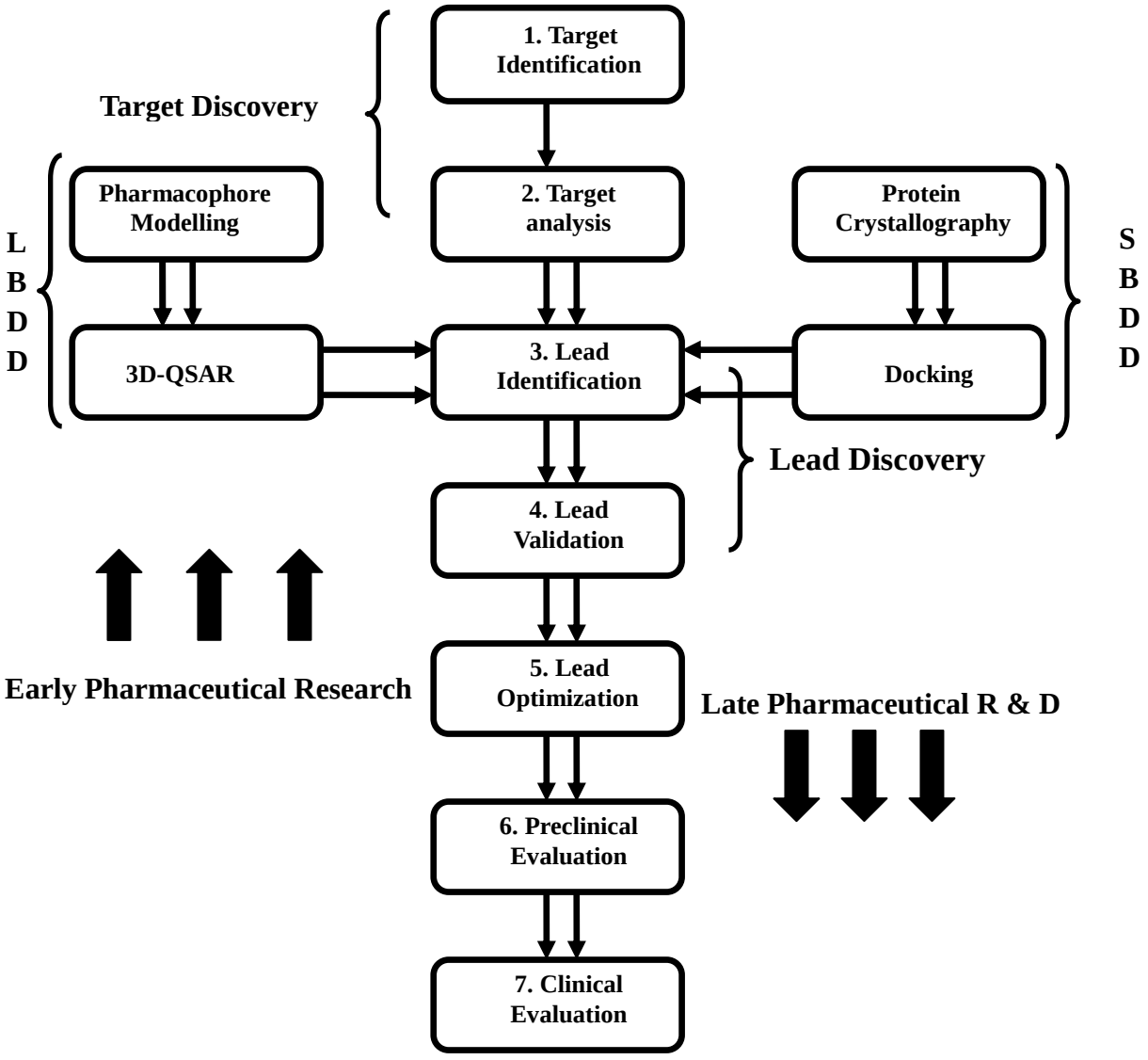

Figure 3: General process of drug design and drug discovery.
LBDD = Ligand based drug design; SBDD = Structure based drug design.

protein in pure form is requirement. It could be done by isolating it from its source, or by cloning its gene into a high expression system. Then after following many steps like crystallization, X-ray data collection, solving the phase problem, model building and final refinement, the final structure is obtained. In NMR technique, the energy levels of atomic nuclei are split up by a magnetic field and transitions between these energy levels are obtained by exciting the sample with radiation whose frequency is equivalent to the energy difference between the two levels. This complete procedure involves the use of pulsed Fourier transform NMR and multidimensional NMR spectroscopy. The NMR techniques for protein structure determination available till date allow determining of the protein structures with masses up to 30 kDa.

In SBDD having the three dimensional protein structure from anyone of the above methods is a requirement. If the structure is not available from any of the above mentioned methods, one can follow the homology modeling to determine the 3D-protein structure to be used for the purpose of SBDD.

Homology modeling, also known as comparative modeling, is a reliable computational tool to determine the 3D structure of proteins whose structures are unknown. The proteins available with the known 3D structures may serve as templates to predict the structure of a protein whose structure is unknown if they share the sequence similarity. Comparative modeling includes the fundamentals of homology modeling in addition to the construction of protein models from templates possessing the same or similar structure (but that may have different biological functions). In this process the important requirements are the protein being modeled (called as the target) and the protein structure(s) that is being used to construct the 3D model (known as template). Overall, a typical protein modeling procedure involves four main steps: (1) finding the template; (2) aligning the target to the template; (3) constructing and refining the protein model; and (4) evaluation of the derived model. Among these, the most difficult steps are alignment of target and template amino acid residues and evaluation of the model. The process of finding the best alignment, construction, refinement and evaluation of a protein model is less of a precise science and more of an art.

Although the homology modeling method seems to be the most reliable one, it can be applied only when 3D structure of a similar sequence or family member is already known.

Threading is another technique which provides an account of the possibility when the sequence finds very poor similarity in one template. In this method the amino acids' sequence is compared with a set of protein structures (i.e. structural patterns); the sequence where it is most likely to fold is computed and pseudoprotein models are constructed. The calculations are performed to determine the energy values for pseudo-models using empirical energy functions and the pseudo-models are ranked on the basis of the energy. The template with the lowest energy value along with a higher percentage of identity or similarity is considered as the most probable template.[30]

1.2.2. Molecular Docking

In the field of molecular modeling, docking is a method which predicts the preferred orientation of one molecule to a second one when bound to each other to form a stable complex. Knowledge of the preferred orientation in turn may be used to predict the strength of association or binding affinity between the two molecules using, for example, scoring functions. Docking is frequently used to predict the binding orientation of small molecule drug candidates to their protein targets in order to predict the affinity and activity of the small molecule drugs. Hence docking plays an important role in the rational design of drugs.[31]

To perform a docking screen, the first requirement is the structure of a protein of interest. Usually the structure has been determined using biophysical techniques such as X-ray crystallography or NMR spectroscopy. This protein structure and a database of potential ligands serve as the inputs to a docking program. The success of a docking program depends on two components- the search algorithm and the scoring function.[32,33]

Search algorithm is nothing but searching the conformational space for docking. The search space in theory consists of all possible orientations and conformations of the protein paired with the ligand. However in practice with current computational resources, it is impossible to exhaustively explore the search space—this would involve enumerating all possible distortions of each molecule (molecules are dynamic and exist in an ensemble of conformational states) and all possible rotational and translational orientations of the ligand relative to the protein at a given level of granularity. Most docking programs utilize flexible ligands, and attempt to model a partially flexible protein receptor. Each "snapshot" of the pair is referred to as a pose. A variety of conformational search strategies have been applied to the ligand and to the receptor. These include: Systematic or stochastic torsion which searches

conformations about rotatable bonds; Molecular dynamics simulations and Genetic algorithms. The scoring function (force field) takes a pose as input and returns a number indicating the likelihood that the pose represents a favorable binding interaction. Scoring functions are physics-based molecular mechanics force fields that estimate the energy of the pose; a low (negative) energy indicates a stable system and thus a likely binding interaction.[34,35]

1.2.3. Molecular Simulation

Simulation is the imitation of the operation of a process or system over a period of time. The act of simulating something first requires that a model be developed; this model represents the key characteristics or behaviors/functions of the selected physical or abstract system or process. The model represents the system itself, whereas the simulation represents the operation of the system over time. A computer simulation is an attempt to model a real-life or hypothetical situation on a computer so that it can be studied to see how the system works over a period of time. By changing variables in the simulation, predictions can be made about the behavior of the system. In chemistry, simulation studies are performed to obtain all the information about possible energy levels, conformations, and so on.[38] The molecular simulations are of two types,

- Monte Carlo (MC) Simulation.
- Molecular Dynamics (MD) Simulation.

MC and MD differ in varieties of ways.

MC is probabilistic approach that relies on repeated random sampling to obtain numerical results. For MC, a new conformation is generated by selecting a random structure of the molecule, translating it, rotating it, and performing any internal structural variations. MD is a simulation of the time-dependent behavior of a molecular system. For MD, new conformations are generated by application of Newton's equations of motion.

MD provides information about the time dependence of the properties of the system whereas there is no temporal relationship between successive MC configurations. In MC simulation the outcome of each trial move depends only upon its immediate predecessor, whereas in MD it is possible to predict the configuration of system at any time in the future or indeed at any time in the past. MD has a kinetic energy contribution to the total energy whereas

in MC simulation the total energy is determined from the potential energy function. In MC one atom is moved at a time and compared with the rest, whereas, in MD all atoms are considered simultaneously over a small time step to determine the new atomic positions and velocities. MC sampling yield 10 times less atomic diffusion than MD for a given amount of computer time in simulations.[36]

1.2.4. Force Field

In the context of molecular modeling, a force field refers to the form and parameters of mathematical functions used to describe the potential energy of a system of particles (typically molecules and atoms). Force field functions and parameter sets are derived from both experimental work and high-level quantum mechanical calculations. "All-atom" force fields provide parameters for every type of atom in a system, including hydrogen, while "united-atom" force fields treat the hydrogen and carbon atoms in each terminal methyl and each methylene bridge as a single interaction center. "Coarse-grained" force fields, which are frequently used in long-time simulations of proteins, provide even more crude representations for increased computational efficiency.[36]

The basic functional form of a force field encapsulates both bonded terms relating to atoms that are linked by covalent bonds, and nonbonded (also called "noncovalent") terms describing the long-range electrostatic and van der Waals forces. The specific decomposition of the terms depends on the force field, but a general form for the total energy in an additive force field can be written as,

$$E_{total} = E_{bonded} + E_{nonbonded}$$

Where the components of the covalent and noncovalent contributions are given by the following summations:

$$E_{bonded} = E_{bond} + E_{angle} + E_{dihedral}$$

$$E_{nonbonded} = E_{elctrostatic} + E_{van\ der\ Waals}$$

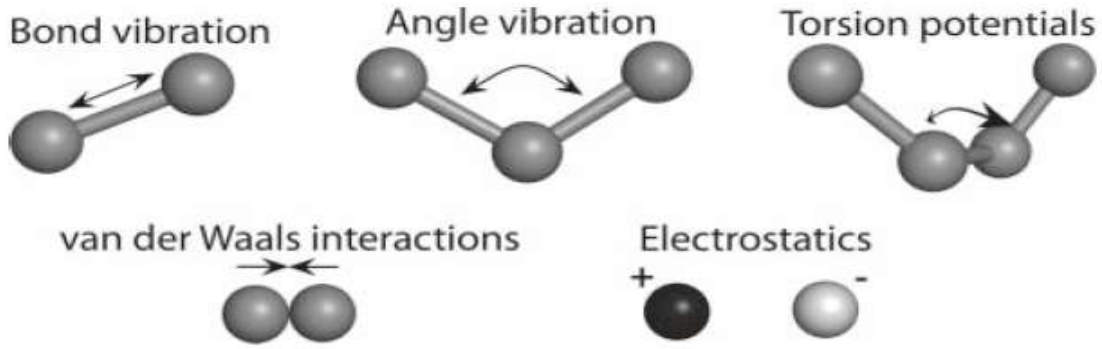

Figure 4: Bonded and non-bonded terms of force field.

1.2.5. Pharmacophore modeling

As per the IUPAC (1998) "A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response."[37]

In medicinal chemistry most of the time this term is used incorrectly. A pharmacophore never represent a real molecule/skeleton such as flavones, steroids etc or never present a real association of functional groups like hydroxyl, amine, guanidines etc, but it is only a summarized concept that accounts for the common molecular interaction capacities of a group of compounds towards their target structure. Pharmacophore modeling helps to identify a set of common features that interacts with a set of complementary sites on the biological target. Generally these features are, hydrogen bond donors, hydrogen bond acceptors, negatively and positively charged groups and hydrophobic regions. Pharmacophore concept is closely related with the concept of bioisosterism. In addition, in the 3D pharmacophore methods the spatial relationship between the pharmacophoric features is also precise. These pharmacophoric features may be positioned on the ligand itself or may be presumed as counter features to be located in the receptor.

In general the process of pharmacophore model development involves five steps. Selection of data set as training set should include the structurally diverse set of molecules. Also the data set should include both active and inactive compounds in order to train the model to discriminate between active and inactive compounds. The second step involves the development of low energy conformations that are likely to contain the biologically active conformations of

individual ligands. The alignment of different conformations generated for ligands either considering atom to atom pairing or feature based pairing is the next step for the pharmacophore model development. Here a set of conformations (one for each ligand) which show good fitting or alignment is presumed to be the most active conformation. Next step is abstraction or representation of the molecules with the common identified features such as aromatic ring, hydrophobic feature, hydrogen bond donor/acceptor etc. And the final step includes the validation of the developed model. One of the basic validation steps is that the model should differentiate the molecules with high and low biological activity. One can also validate the generated pharmacophore model by comparing the mapped features with the active site of the receptor under study, if the receptor 3D structure is available. It is also useful to compare the features of the pharmacophore with the features obtained from contours of nD-QSAR.[38,39]

Apart from this traditional concept of pharmacophore, one can generate receptor dependent pharmacophore i.e. mapping of pharmacophore on the receptor active site.

1.2.6. Virtual Screening

Virtual screening (VS)[40] is a computational technique used in the drug design process to screen a library of small molecules in order to identify the probable active structures which are most likely to bind with the active site of the target receptor or enzyme. VS can search enormous chemical space of over 10^{60} plausible compounds to a handy number that can be synthesized and evaluated against a particular condition. Although searching in of such a large data may be interesting theoretically, but in practical terms VS focuses more on selective or targeted combinatorial libraries.

The VS technique can be either structure based or ligand based. Structure-based VS involves docking of candidate ligands into the active site of the target protein or receptor and then applying a scoring function. The ligands that show high affinity for the active site of the target receptor are identified.

The ligand based or pharmacophore based VS or logic-based rules describe features of substructures and chemical properties related to the activity using support vector inductive logic programming. Another approach is molecular similarity searching which is based upon similar property principle and attempts to predict properties of molecules on the basis of knowledge

derived from properties of another set of molecules. The chemical similarity study is useful to compare the biological similarity between a set of molecules.

1.2.7. Quantitative Structure Activity Relationship

Quantitative structure activity relationship (QSAR)[11] or quantitative structure property relationship (QSPR), is the process in which chemical structure is quantitatively correlated with some well defined process, such as biological activity or chemical reactivity. When physicochemical properties or structures are expressed with numbers, one can form a mathematical relation between the structure and the activity. The mathematical expression can then be used to predict the biological response of other chemical structures. QSAR is based on the assumption that there is an underlying relationship between the molecular structure and biological activity. On this assumption QSAR attempts to establish a correlation between various molecular properties of a set of molecules with their experimentally determined biological activity. There are two main objectives for the development of QSAR:

- Development of predictive and robust QSAR, with a specified chemical domain, for the prediction of activity of untested molecules.
- It acts as an informative tool by extracting significant patterns in descriptors related to the measured biological activity leading to understanding the mechanism of the biological activity. This could help in the designing of novel molecules with improved activity profile.

A general mathematical equation for QSAR is:

Activity = *f* (physicochemical and/or structural properties)

For the generation of QSAR or to perform QSAR study the general method involves the following steps:

For QSAR analysis, a set of series of synthesized molecules tested for their desired activity is required. Its quality totally depends on the quality of the experimental data used for building the model. The biological activity used is in terms of mathematical data obtained by applying some statistics. This can be of two type; continuous response type: such as Minimum effective concentration (MEC), Inhibitory concentration (IC50), Effective dose (ED50), % inhibition, and categorical response such as active or inactive etc.

Molecular descriptors can be defined as numerical representation of chemical information encoded with molecular structure via some mathematical procedure. There are so many descriptors available for building a QSAR model, but all are not important for a specific model generation, and hence to select an optimal set of descriptors, various selection methods are used which include systematic variable selection and stochastic variable selection. In systematic variable selection leave one out rule is used and it is applied by stepwise forward, stepwise forward-backward or by stepwise backward methods. Stochastic variable selection methods are based on simulation of various physical or biological processes. These methods create models starting from randomly generated models and later modifying these models by using different process operators, such as perturbation, crossover etc to get better model(s). The descriptors used in QSAR are based on the dimensionality of the molecule. For example from zero dimensional molecule, molecular weight, number and type of atoms etc are calculated; from one dimensional molecular representation functional groups, ring, bonds etc are calculated; from two dimensional, types of bonding and interaction of particular atoms is possible; from three dimensional molecular representation various descriptors, such as molecular surface, molecular volume, electronic, steric and geometric descriptors are derived. Apart from these dimensions one can consider the fourth dimension, generally in terms of conformational flexibility or the fifth dimension in terms of additional different induced fit features.

In the QSAR study the model need to be validated and for this purpose the dataset is to be divided into training set (for building the QSAR model) and test set (for examining its predictive ability). For any QSAR model, it is of importance that the training set selected to create a model exhibits a well balanced distribution of biological activity and contains representative molecules. Various methods such as manual selection method, random selection method, sphere exclusion method, factorial design, cluster analysis method etc are used for the division of dataset into training and test set.

The selected variable descriptors when coupled with a suitable statistical method allow analyses of data subsets in order to establish a QSAR model for the selected descriptors. Such model should be statistically significant in determining the biological activity. The statistical methods used in QSAR can be broadly divided into linear and non-linear methods. In statistics a correlation is established between dependent variable(s) (biological activity) and independent variable(s) (molecular descriptors). The linear method fits a line between the selected descriptors

and activity as compared to non-linear method which fits a curve between the selected descriptors and activity. The statistical method to build the QSAR model is decided on the basis of type of biological activity data. Various statistical methods includes discriminant analysis, k-nearest neighbor classification, multiple regression, principle component regression, continuum regression, partial least square analysis etc. which can be used for specific set of data.

The procedure followed for the development of 3D- and 4D-QSAR models on aurora A kinase inhibitors along with the results and discussion is described in detail here in section I.